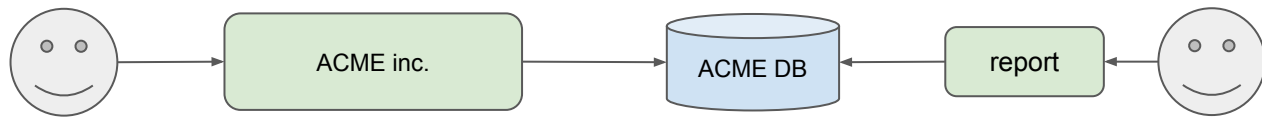
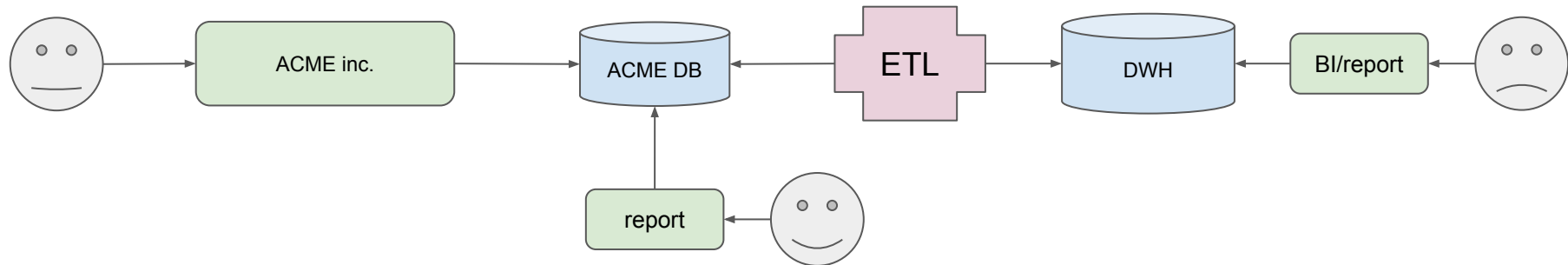


Prospettive per data pipeline

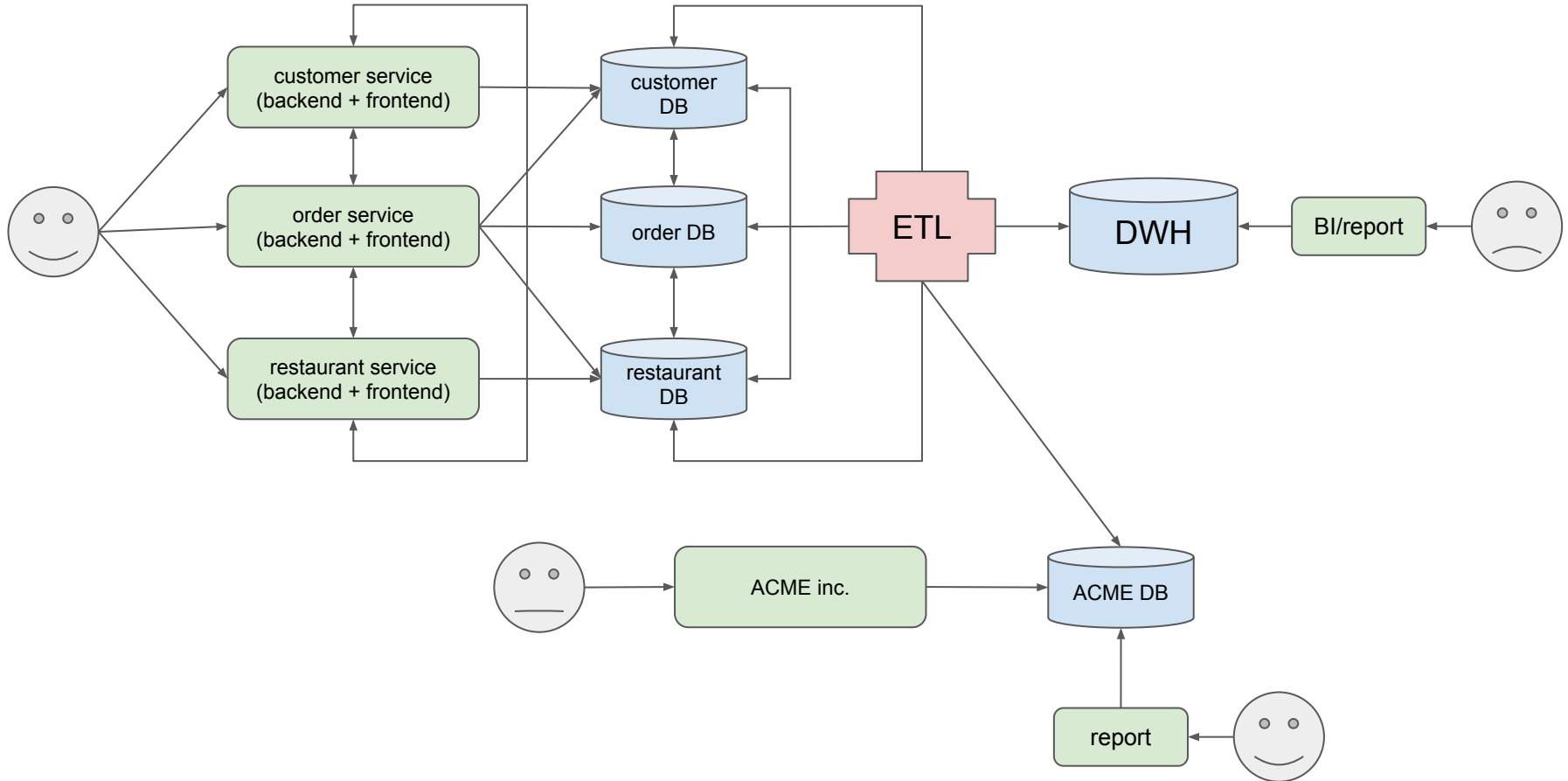
C'era una volta.....



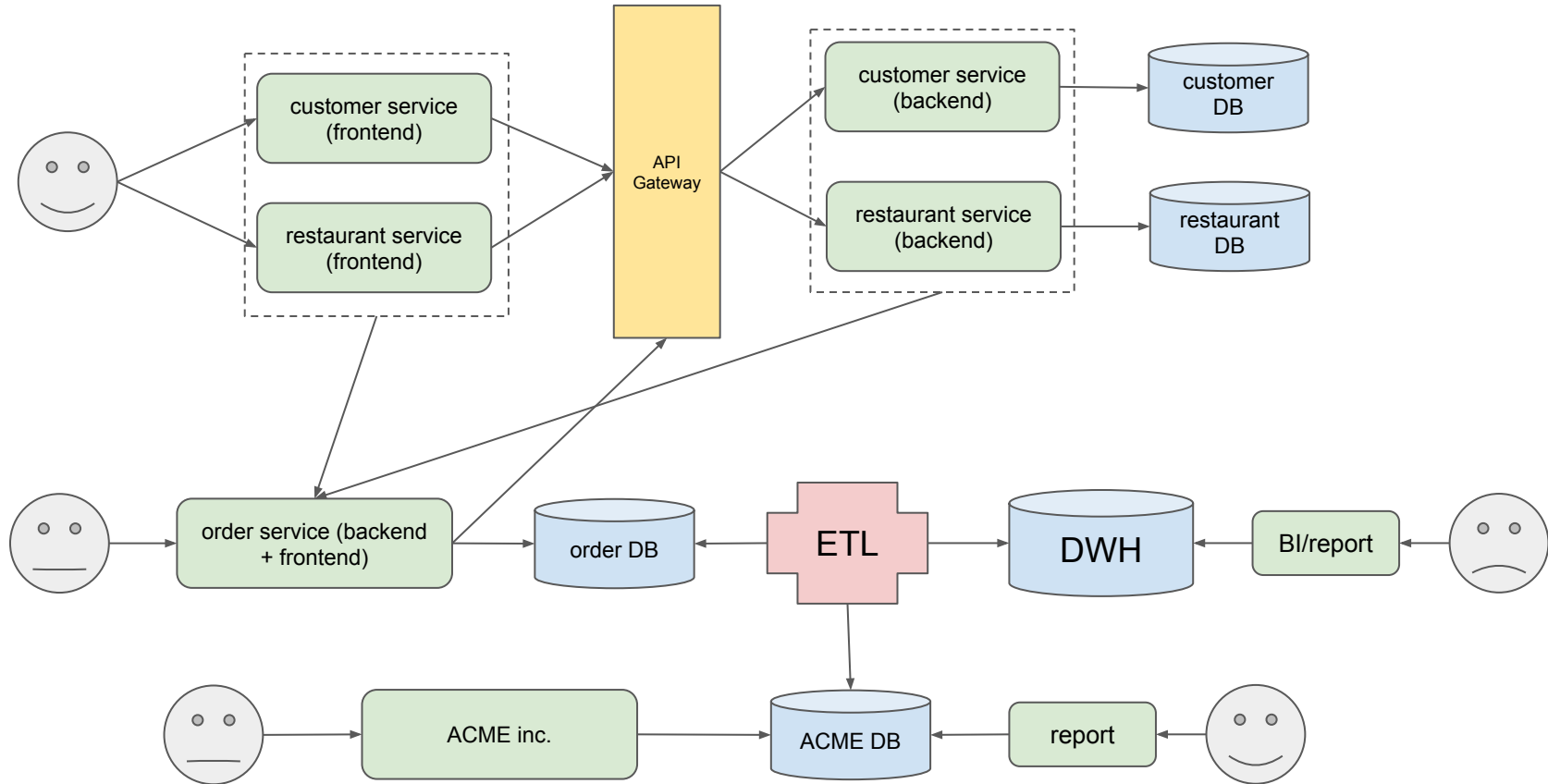
introduzione DWH



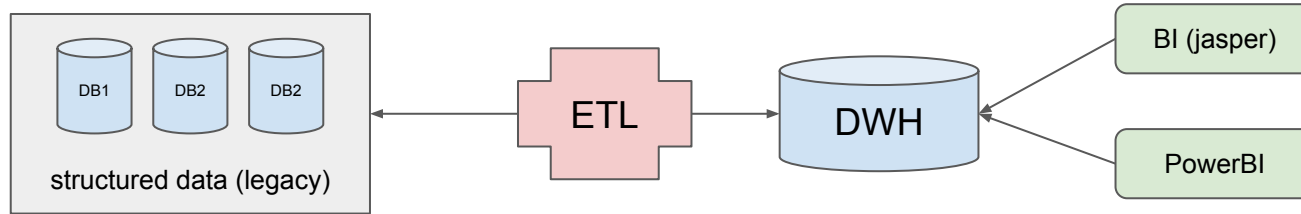
Che è degenerato in...



Che ad oggi è così (dopo adozione microservizi)



Flusso “legacy”: problemi e limitazioni

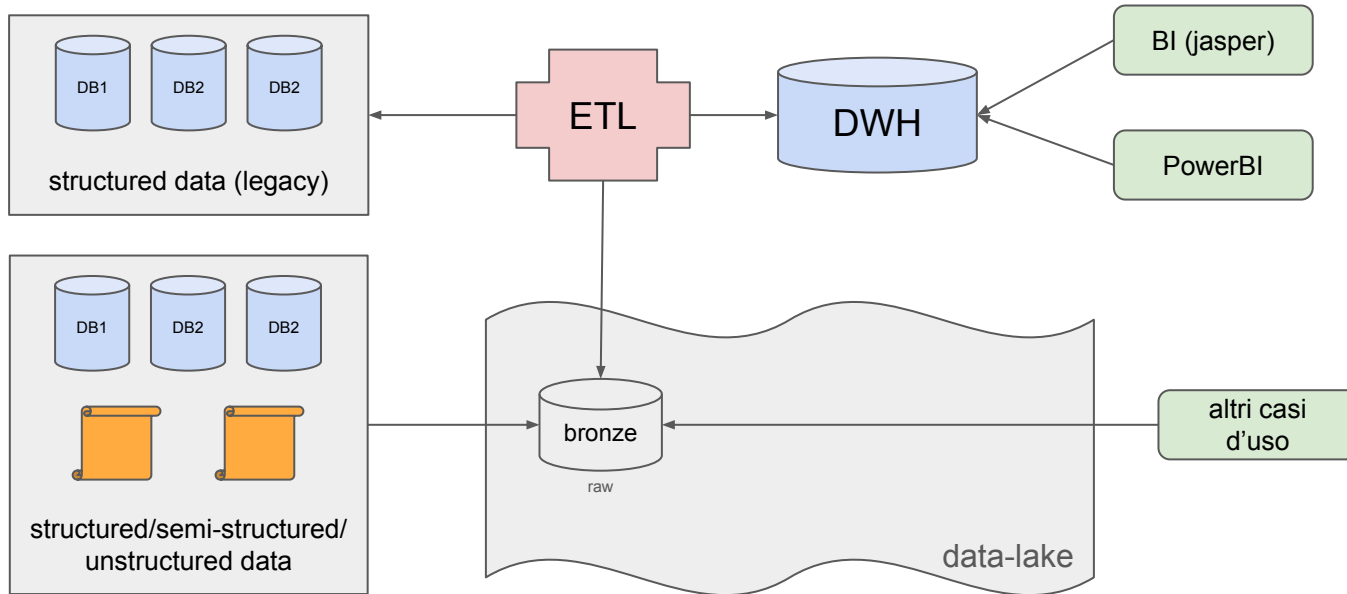


- pensato per dati strutturati, ambienti monolitici e per casi d'uso molto semplici/applicazioni legacy
- authN/authZ non ottimale
- accesso diretto al dato grezzo + grosso impatto su database sorgenti e su DWH
- refactoring struttura dati complicato (es: cambio DB engine, cambio schema dati)
- inefficienza dei job (es: tipicamente leggono tutto ad ogni esecuzione) + tempi lunghi di esecuzione
- collaborazione/versioning e altre pratiche di sviluppo software difficilmente applicabili
- costi di gestione alti (setup non ottimizzato per questo caso d'uso, ambiente prone ad errori)
- ETL totalmente a carico di un gruppo ristretto di persone “specializzate”:
 - che non sanno (e non devono sapere) nulla (o quasi) sulla struttura grezza dato sorgente normalizzato
 - che devono pre-lavorare il dato grezzo con logiche che non conoscono
 - richiede over coordinamento con chi progetta le applicazioni
 - comportamento che porta a “de-responsabilizzazione” su correttezza (forma e contenuto) del dato prodotto dalle applicazioni (che non deve essere fine a se stesso)

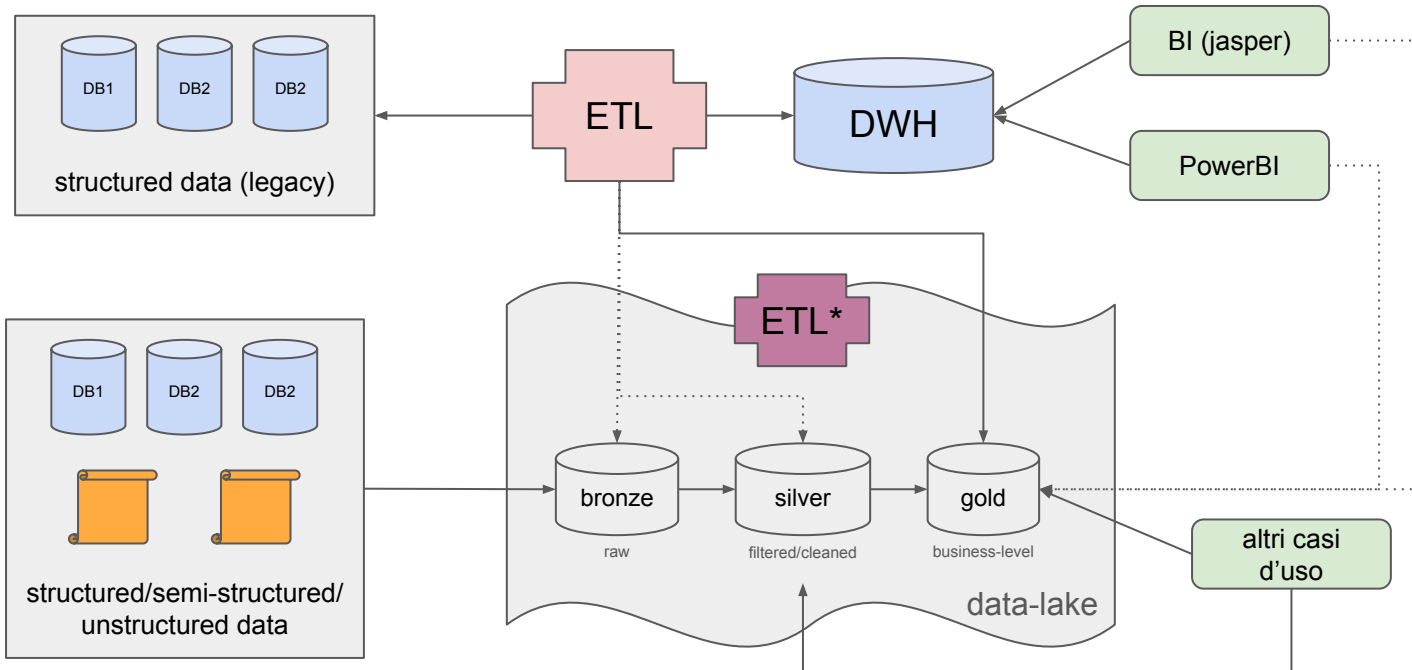
Cosa vogliamo

- fornire una soluzione/piattaforma che si possa adattare alle esigenze odierne ma che permetta evoluzioni future
 - rendere fruibili i dati presenti negli applicativi di “nuova concezione”
 - poter ricevere dati da fonti non omogenee
 - poter utilizzare il dato per casi d’uso diversi dalla “semplice” reportistica
- ottimizzare e migliorare i processi permettendo anche una transizione “soft” (o un parallelo completo)
- migliorare le modalità di lavoro su tale impianto (es: pratiche di CI/CD e test per i job ETL)
- non sovraccaricare di lavoro più del dovuto chi sviluppa software o chi si occupa di BI
- risolvere i problemi tecnici attualmente presenti (con soluzioni più o meno pronte) senza scendere a grossi compromessi
- de-centralizzare la responsabilità sui dati
- migliorare governance sui dati

Possibile evoluzione: step 1

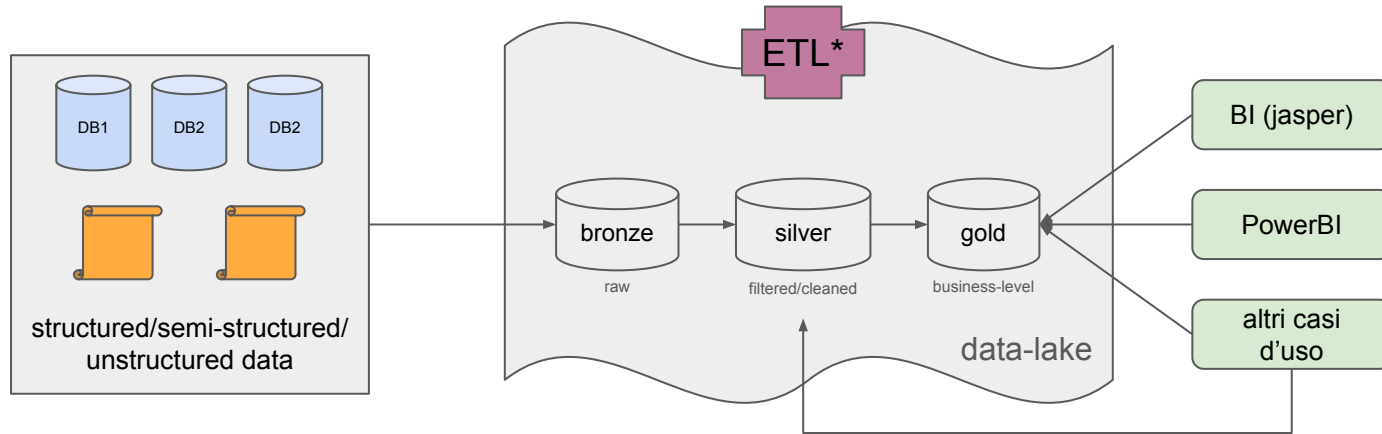


Possibile evoluzione: step 2



(*) meglio se qualcosa di più specifico, evoluto e moderno dell'attuale

Possibile stato finale



Altre opzioni

- non cambiamo nulla e ci inventiamo un alternativa, es:
 - ETL legge direttamente da sorgente per i microservizi
 - per mongo, export notturno su s3
 - altro?
- altro?